



Capacities of the
Mean-Square-Constrained Poisson Channel

Michael R. Frey

Department of Statistics
University of North Carolina
Chapel Hill, NC 27599

LISS 45

4 July 1990

DTIC
ELECTE
MAY 13 1991
S E D

Abstract - An earlier discussion of the capacity of the mean-square-constrained Poisson channel is continued. Using a theorem of Hoeffding, it is shown that the channel information capacity is the same with or without an on-off keying (OOK) constraint on the channel encoder intensity, affirmatively resolving the conjecture made in our earlier discussion. Thus the known formula for the information capacity of the OOK-constrained channel applies as well in the absence of an OOK constraint. Adapting arguments used by Wyner to address the peak-constrained Poisson channel, it is also shown that the coding capacity with no OOK constraint is equal to the corresponding information capacity. This establishes that all four capacities - coding and information, with and without OOK-constrained encoder - are equal.



Supported by ARO Contract DAAL03-86-G-0022, ONR Contract N00014-86-K-0039 and NSF Contract NCR8713726.

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

91 5 13 068

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release: Distribution Unlimited		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S)			5. MONITORING ORGANIZATION REPORT NUMBER(S)		
6a. NAME OF PERFORMING ORGANIZATION Department of Statistics		6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION		
6c. ADDRESS (City, State and ZIP Code) University of North Carolina Chapel Hill, North Carolina 27514			7b. ADDRESS (City, State and ZIP Code)		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research		8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0039		
8c. ADDRESS (City, State and ZIP Code) Statistics & Probability Program Arlington, VA 22217			10. SOURCE OF FUNDING NOS.		
11. TITLE (Include Security Classification) Capacities of the Mean-Square-Constrained *			PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.
12. PERSONAL AUTHOR(S) M.R. Frey			WORK UNIT NO.		
13a. TYPE OF REPORT TECHNICAL		13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day) 1990 July		15. PAGE COUNT 12
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB. GR.	Poisson channel, Shannon theory, channel capacity		
19. ABSTRACT (Continue on reverse if necessary and identify by block number)					
<p>An earlier discussion of the capacity of the mean-square-constrained Poisson channel is continued. Using a theorem of Hoeffding, it is shown that the channel information capacity is the same with or without an on-off keying (OOK) constraint on the channel encoder intensity, affirmatively resolving the conjecture made in our earlier discussion. Thus the known formula for the information capacity of the OOK-constrained channel applies as well in the absence of an OOK constraint. Adapting arguments used by Wyner to address the peak-constrained Poisson channel, it is also shown that the coding capacity with no OOK constraint is equal to the corresponding information capacity. This establishes that all four capacities - coding and information, with and without OOK-constrained encoder - are equal.</p> <p>TITLE CONT.: Poisson Channel</p>					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input checked="" type="checkbox"/> DTIC USERS <input type="checkbox"/>			21. ABSTRACT SECURITY CLASSIFICATION		
22a. NAME OF RESPONSIBLE INDIVIDUAL C.R. Baker			22b. TELEPHONE NUMBER (Include Area Code) (919) 962-2189	22c. OFFICE SYMBOL	

INTRODUCTION

We continue the discussion of the capacity of the mean-square-constrained Poisson channel with causal feedback and general channel base measure from [5]. The Poisson channel is a continuous-time additive noise model. In the model, the channel output Y is a stochastic process $Y_t = X_t + N_t$, where $N = \{N_t\}_{0 \leq t \leq T}$ is the channel noise and $X = \{X_t\}_{0 \leq t \leq T}$ is the transmitted signal into which is encoded the message $\theta = \{\theta_t\}_{0 \leq t \leq T}$ (assumed to be independent of the noise process N .) Here, X and N are Poisson-type point processes [8] with respective compensators

$$A_t = \int_0^t \chi_s b(ds), \quad B_t = \int_0^t \lambda b(ds) = \lambda b([0, t])$$

for all $t \in [0, T]$. b is called the channel base measure and is assumed to be finite; for convenience we assume $b_T = 1$ where $b_T \equiv b([0, T])/T$. Also, the noise intensity $\lambda \geq 0$ is considered as a real constant. A mean-square constraint $E[\chi_t^2] \leq P^2$ is imposed on the encoder intensity χ_t . χ_t is also restricted to be predictable, allowing only causal feedback. For more on the predictability restriction and mean-square constraint and for further general description and discussion of the Poisson channel model, the reader is invited to review [5].

A special case of both practical and theoretical interest is that in which the encoder intensity χ_t is constrained to switch between only two values. In this situation an "on-off" keying (OOK) constraint is said to be imposed on the encoder intensity in recognition of the fact that when the range of χ_t is restricted to just two values, then one of these values should be zero (to minimize the effective channel noise intensity.) In an actual implementation of such an encoder, transitions of χ_t between its zero value and its second (positive) value might typically be accomplished by turning on and off a power source. Hence the nomenclature "on-off keying". We consider the Poisson channel with and without an OOK constraint on the encoder intensity.

Information capacity is defined in terms of the average mutual information $I^T[\theta, Y]$ in the message and channel output processes, θ and Y over the interval $[0, T]$. Let μ_θ , μ_Y , and $\mu_{\theta Y}$ be the marginal and joint measures induced by the message and output processes, θ and Y , on the spaces S_θ , S_Y , and $S_\theta \times S_Y$ where S_θ and S_Y are the spaces of trajectories of θ and Y over the interval $[0, T]$. Write the induced product measure as $\mu_{\theta \times Y}$. Then, the average mutual information in θ and Y over the interval $[0, T]$ is [9]

$$I^T[\theta, Y] = E \left[\ln \frac{d\mu_{\theta Y}}{d\mu_{\theta \times Y}} \right]$$

provided $\mu_{\theta Y} \ll \mu_{\theta \times Y}$; otherwise $I^T[\theta, Y] = \infty$. Expressions exist for the average mutual information over the interval $[0, T]$ in the Poisson channel with base measure b and channel output intensity η_t . Under the conditions of the channel model

$$I^T[\theta, Y] = E \left[\int_0^T (\eta_t \ln \eta_t - \hat{\eta}_t \ln \hat{\eta}_t) b(dt) \right] \quad (1)$$

where $\eta_t = \chi_t + \lambda$ is the channel output intensity and $\hat{\eta}_t$ is the predictable version of the mean of η_t , conditioned on the path of Y up to time t . The channel information capacity is

$$C_{\text{INFO}} = \sup_{\theta} \sup_{\chi} \frac{1}{T} I^T[\theta, Y]$$

where θ is any jointly measurable process defined over the interval $[0, T]$ and $\chi = \chi(\theta, Y)$ is any predictable functional satisfying the mean-square constraint $E[\chi_t^2] \leq P^2$. We write $C_{\text{INFO}}^{\text{OOK}}$ and C_{INFO} , respectively, for the information capacities with and without OOK-constrained encoder intensity. To emphasize the

dependence of the information capacity on the channel parameters λ and P , we use the notation

$$C_{\text{INFO}} = D(\lambda, P), \quad C_{\text{INFO}}^{\text{OOK}} = D_o(\lambda, P).$$

The functions $D(\lambda, P)$ and $D_o(\lambda, P)$ are easily shown [5] to be first-order homogeneous; i.e., $D(z\lambda, zP) = zD(\lambda, P)$ and $D_o(z\lambda, zP) = zD_o(\lambda, P)$. Thus to generally determine the information capacities C_{INFO} and $C_{\text{INFO}}^{\text{OOK}}$, one need only find expressions for

$$D(0, P), \quad D_o(0, P), \quad D(1, P), \quad D_o(1, P).$$

In our initial treatment [5] of the Poisson channel with a mean-square constraint, a simple formula was obtained for the information capacity for the special case of zero noise intensity. There it was shown that

$$D(0, P) = D_o(0, P) = \frac{2}{e} P.$$

The case of nonzero noise intensity was found to be less tractable and an explicit expression for the information capacity was found only for the special case in which the encoder intensity is OOK-constrained:

$$D_o(1, P) = \sup_{\alpha^2 k(1/\alpha) \leq P^2} \frac{P^2}{\alpha} \ln \frac{\alpha k(1/\alpha) + 1}{P^2 k(\alpha/P^2)/\alpha + 1} \quad (2)$$

where

$$k(x) = \frac{x}{e} \left(1 + \frac{1}{x} \right)^{x+1} - x.$$

The remainder of the paper is comprised of three sections. The first uses a theorem of Hoeffding [7] to show that $D(1, P) = D_o(1, P)$, allowing (2) to be used to calculate the capacity $C_{\text{INFO}} = D(\lambda, P)$. The second section introduces coding capacity and shows it to be equal to the information capacity. Some final remarks are made in the last section.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input checked="" type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	



INFORMATION CAPACITY

The purpose of this section is to show that the information capacities C_{INFO} and $C_{\text{INFO}}^{\text{OOK}}$ are equal. This is known to be true in the special case $\lambda=0$ of zero noise intensity. Thus, showing $C_{\text{INFO}} = C_{\text{INFO}}^{\text{OOK}}$ amounts to showing

$$D(1, P) = D_o(1, P). \quad (3)$$

(3) is proved in part by an application of a theorem of Hoeffding [7] on the extrema of expectations. We prepare the way for the use of this theorem with some notation and several lemmas.

Let $\mathcal{B}(P)$ denote the class of nonnegative random variables X satisfying $E[X^2] = P^2$. Also let $\mathcal{B}^*(P) \subset \mathcal{B}(P)$ denote those members X of $\mathcal{B}(P)$ which are discrete. We use $\mathcal{B}_k(P)$ to denote the class of $X \in \mathcal{B}^*(P)$ with no more than k atoms. We also freely use $\mathcal{B}(P)$, $\mathcal{B}^*(P)$, and $\mathcal{B}_k(P)$ to denote, respectively, the classes of distribution functions generated by $X \in \mathcal{B}(P)$, $X \in \mathcal{B}^*(P)$, and $X \in \mathcal{B}_k(P)$. The distance $d(F, G)$ between two distribution functions $F, G \in \mathcal{B}(P)$ is defined to be

$$d(F, G) = \sup_{x \geq 0} |F(x) - G(x)|.$$

We also define the functional $\phi(F)$ for $F \in \mathcal{B}(P)$ by

$$\phi(F) = \int_0^\infty (x+1) \ln(x+1) F(dx).$$

$\phi(F)$ exists and is finite for all $F \in \mathcal{B}(P)$. Indeed, $\phi(F) \leq P^2 + P$ for all $F \in \mathcal{B}(P)$.

Lemma 1: For every $\varepsilon > 0$ and $F \in \mathcal{B}(P)$ there exists a distribution function $F^* \in \mathcal{B}^*(P)$ such that

$$d(F, F^*) < \varepsilon. \quad (4)$$

Proof: (4) holds for $F^*(x) = \lfloor \frac{1}{\varepsilon} \rfloor F(x + \Delta) / \lfloor \frac{1}{\varepsilon} \rfloor$ for some choice of $\Delta > 0$.

Lemma 2: For every $\varepsilon > 0$ and $F \in \mathcal{B}(P)$, there exists a $\delta > 0$ such that

$$|\phi(F) - \phi(G)| < \varepsilon$$

for all $G \in \mathcal{B}(P)$ satisfying $d(F, G) < \delta$.

Proof: Let $F \in \mathcal{B}(P)$. Let $G \in \mathcal{B}(P)$ be any distribution function such that $d(F, G) < \delta$. Associate with F and G the truncated distribution functions

$$F_B(x) = \begin{cases} F(x), & x < B \\ 1, & x \geq B \end{cases},$$

$$G_B(x) = \begin{cases} G(x), & x < B \\ 1, & x \geq B \end{cases}.$$

The idea for the proof is to show that $\phi(F)$ and $\phi(F_B)$ differ by an arbitrarily small amount for a large enough choice of B . Likewise for $\phi(G)$ and $\phi(G_B)$. Next, F_B and G_B are approximated by discrete distribution functions F^n and G^n with n equal size jumps. We show that the differences $|\phi(F_B) - \phi(F^n)|$ and $|\phi(G_B) - \phi(G^n)|$ can be made arbitrarily small by improving (letting $n \rightarrow \infty$) the approximations of F_B , G_B by F^n , G^n . Finally, to complete the proof, it is shown that, for $\delta \rightarrow 0$, $|\phi(F^n) - \phi(G^n)|$ is

arbitrarily small uniformly in G for all G within distance δ of F .

We have

$$\begin{aligned} |\phi(F) - \phi(G)| &\leq |\phi(F) - \phi(F_B)| + |\phi(G) - \phi(G_B)| + |\phi(F_B) - \phi(F^n)| \\ &\quad + |\phi(G_B) - \phi(G^n)| + |\phi(G^n) - \phi(F^n)| \end{aligned}$$

where, for positive integers n , F^n and G^n are discrete approximations of F_B and G_B ; F^n and G^n are, respectively, the right-continuous versions of the functions

$$\frac{1}{n} [nF(\cdot)], \quad \frac{1}{n} [nG(\cdot)].$$

We have

$$|\phi(F) - \phi(F_B)| \leq \int_B^\infty (x+1) \ln(x+1) F(dx) - (B+1) \ln(B+1) (1 - F(B)). \quad (5)$$

The integral on the RHS of (5) is finite so B can be chosen so that

$$\int_B^\infty (x+1) \ln(x+1) F(dx) < \varepsilon_1 \quad (6)$$

for any $\varepsilon_1 > 0$. Now

$$0 \leq (B+1) \ln(B+1) (1 - F(B)) \leq \int_B^\infty (x+1) \ln(x+1) F(dx),$$

so

$$|\phi(F) - \phi(F_B)| \leq \varepsilon_1$$

for B large enough. We choose B so that (6) and a similar statement for G hold. Then

$$\begin{aligned} |\phi(F) - \phi(G)| &\leq |\phi(F_B) - \phi(F^n)| + |\phi(G_B) - \phi(G^n)| \\ &\quad + |\phi(G^n) - \phi(F^n)| + 2\varepsilon_1. \end{aligned}$$

Define F_n to be the distribution function $F^n(x) = [nF(x)]/n$. Let $\{x_i, i=0, 1, \dots, n\}$ be the locations of the jumps of F_n and F^n ; the jumps of F^n are located at $\{x_i, i=0, 1, \dots, n-1\}$ and the jumps of F_n are located at $\{x_i, i=1, 2, \dots, n\}$. Then

$$\begin{aligned} |\phi(F_B) - \phi(F^n)| &\leq \phi(F_n) - \phi(F^n) \\ &= \frac{1}{n} \sum_{i=1}^n (x_i + 1) \ln(x_i + 1) - \frac{1}{n} \sum_{i=0}^{n-1} (x_i + 1) \ln(x_i + 1) \\ &= \frac{1}{n} (x_n + 1) \ln(x_n + 1) - \frac{1}{n} (x_0 + 1) \ln(x_0 + 1) \\ &\leq \frac{1}{n} (B + 1) \ln(B + 1). \end{aligned}$$

Consider $n \geq [(B+1) \ln(B+1) / \varepsilon_2]$. For any such n and given any $\varepsilon_2 > 0$, we have $|\phi(F_B) - \phi(F^n)| \leq \varepsilon_2$. Given our choice of B , the same n gives $|\phi(G_B) - \phi(G^n)| \leq \varepsilon_2$. Therefore

$$|\phi(F) - \phi(G)| \leq |\phi(G^n) - \phi(F^n)| + 2\varepsilon_1 + 2\varepsilon_2$$

for sufficiently large n and B . For $d(F, G) < \delta$,

$$|\phi(G^n) - \phi(F^n)| \leq \phi(F_n^\delta) - \phi(F^n)$$

where $F_n^\delta(x) = F_n(x - \delta)$. Thus

$$|\phi(G^n) - \phi(F^n)| \leq \frac{1}{n} \sum_{i=1}^n (x_i + \delta + 1) \ln(x_i + \delta + 1) - \frac{1}{n} \sum_{i=0}^{n-1} (x_i + 1) \ln(x_i + 1)$$

Choose δ so that

$$\delta \leq \min_{1 \leq i \leq n} (x_i - x_{i-1})$$

and define $x_{n+1} = x_n + \delta$. Then $x_{i+1} - \delta \geq x_i$ for all $i = 0, 1, \dots, n$. Thus

$$\begin{aligned} |\phi(F_n) - \phi(G_n)| &\leq \frac{1}{n} \sum_{i=1}^n (x_{i+1} + 1) \ln(x_{i+1} + 1) - \frac{1}{n} \sum_{i=0}^{n-1} (x_i + 1) \ln(x_i + 1) \\ &= \frac{1}{n} \left[(x_n + 1) \ln(x_n + 1) + (x_{n+1} + 1) \ln(x_{n+1} + 1) \right. \\ &\quad \left. - (x_0 + 1) \ln(x_0 + 1) - (x_1 + 1) \ln(x_1 + 1) \right] \\ &\leq \frac{2}{n} (B + \delta + 1) \ln(B + \delta + 1). \end{aligned}$$

Let $n = \lceil (B + 2) \ln(B + 2) / \epsilon_3 \rceil$. For sufficiently small $\delta \in (0, 1)$ and all $G \in \mathcal{B}(P)$ satisfying $d(F, G) < \delta$, B and n can be chosen large enough to give

$$|\phi(F) - \phi(G)| \leq 2\epsilon_1 + 4\epsilon_3$$

for any $\epsilon \equiv 2\epsilon_1 + 4\epsilon_3 > 0$. This is what we sought to prove.

We are now in a position to use Hoeffding's theorem.

Lemma 3: Define $I[X] = E[X \ln X] - E[X] \ln E[X]$. Then

$$\sup_{X \in \mathcal{B}(P)} I[X + 1] = \sup_{X \in \mathcal{B}_3(P)} I[X + 1]. \quad (7)$$

Proof: Writing $\mu = E[X]$ and F for the distribution function of X , we have

$$\begin{aligned} \sup_{X \in \mathcal{B}(P)} I[X + 1] &= \sup_{\mu \leq P} \left[\sup_{\substack{X \in \mathcal{B}(P) \\ \mu = E[X]}} \phi(F) - (\mu + 1) \ln(\mu + 1) \right], \\ \sup_{X \in \mathcal{B}_3(P)} I[X + 1] &= \sup_{\mu \leq P} \left[\sup_{\substack{X \in \mathcal{B}_3(P) \\ \mu = E[X]}} \phi(F) - (\mu + 1) \ln(\mu + 1) \right]. \end{aligned}$$

Lemmas 1 and 2 show that ϕ satisfies the conditions required in Hoeffding's theorem [7, Theorem 2.1]. The proof is complete since for constrained first and second moments, Hoeffding's theorem states that

$$\sup_{\substack{X \in \mathcal{B}(P) \\ \mu = E[X]}} \phi(F) = \sup_{\substack{X \in \mathcal{B}_3(P) \\ \mu = E[X]}} \phi(F).$$

Lemma 3 is enough to prove that three-level encoder intensities are sufficient to realize channel information with rate approaching C_{INFO} . Two-level encoding - the on-off keying case - can be shown to be sufficient by inspection of $I[X + 1]$ for $X \in \mathcal{B}_3(P)$. We express $X \in \mathcal{B}_3(P)$ by

$$X = \begin{cases} 0, & \text{w.p. } 1 - p_1 - p_2 \\ a_1, & \text{w.p. } p_1 \\ a_2, & \text{w.p. } p_2 \end{cases}$$

for nonnegative numbers p_1, p_2, a_1, a_2 such that $p_1 + p_2 \leq 1$ and $p_1 a_1^2 + p_2 a_2^2 \leq P^2$. The phrase "with probability" is abbreviated by "w.p." here. Consider the 4-tuple (p_1, p_2, a_1, a_2) as a point $\vec{z} \in \mathbb{E}_4$ belonging to the set

$$A = \{\vec{z} \in \mathbb{E}_4: p_1, p_2, a_1, a_2 \geq 0, p_1 + p_2 \leq 1, p_1 a_1^2 + p_2 a_2^2 \leq P^2\}.$$

$f(\vec{z}) = I[X+1]$ defined on A is continuous. To be closed, the set A lacks only the limit points associated with sequences of points $\{\vec{z}_k = (p_{1k}, p_{2k}, a_{1k}, a_{2k})\}$ with third, fourth, or both coordinates $a_{1k}, a_{2k} \rightarrow \infty$ as $k \rightarrow \infty$. Therefore the supremum of $f(\vec{z})$ over $\vec{z} \in A$ (equivalently, the supremum of $I[X+1]$ over $X \in \mathcal{B}_3(P)$) is to be found as the value of $f(\vec{z})$ for some particular value (or values) of $\vec{z} \in A$ or as the limit of a sequence $\{f(\vec{z}_k)\}$ where $\{\vec{z}_k\}$ is a sequence of points $\vec{z}_k \in A$ with one or both coordinates $a_{1k}, a_{2k} \rightarrow \infty$. We consider three cases exhausting the possibilities for the supremum of $f(\vec{z})$: the case in which at least one coordinate of \vec{z} is zero, the case in which the supremum is the limit with $a_1 \rightarrow \infty$ or $a_2 \rightarrow \infty$, and the case in which no coordinate of \vec{z} is zero.

Let $\vec{z}_0 \in A$ have a zero coordinate and suppose the supremum of $f(\vec{z})$ over A is $f(\vec{z}_0)$. In this case, X actually has only two atoms and two-level encoding suffices to obtain channel information rates approaching capacity.

Suppose the supremum is the limit of a sequence $\{f(\vec{z}_k)\}$ where the coordinate a_{1k} of \vec{z}_k tends to infinity and a_{2k} is bounded. Then

$$\begin{aligned} \lim_{k \rightarrow \infty} f(\vec{z}_k) &= \lim_{k \rightarrow \infty} \left[p_{1k}(a_{1k}+1)\ln(a_{1k}+1) + p_{2k}(a_{2k}+1)\ln(a_{2k}+1) \right. \\ &\quad \left. - (p_{1k}a_{1k} + p_{2k}a_{2k} + 1)\ln(p_{1k}a_{1k} + p_{2k}a_{2k} + 1) \right] \\ &\leq \lim_{k \rightarrow \infty} \left[(P^2/a_{1k}^2)(a_{1k}+1)\ln(a_{1k}+1) + p_{2k}(a_{2k}+1)\ln(a_{2k}+1) \right. \\ &\quad \left. - (p_{2k}a_{2k} + 1)\ln(p_{2k}a_{2k} + 1) \right] \\ &\leq \lim_{k \rightarrow \infty} \left[p_{2k}(a_{2k}+1)\ln(a_{2k}+1) - (p_{2k}a_{2k} + 1)\ln(p_{2k}a_{2k} + 1) \right]. \end{aligned}$$

This shows the supremum to be dominated by the limit of a sequence of values of $I[X_k+1]$ where $X_k \in \mathcal{B}_2(P)$. The symmetric case in which $a_{2k} \rightarrow \infty$ while a_{1k} is bounded and the case where both coordinates tend to infinity are handled similarly.

Finally, consider the case in which the supremum of $f(\vec{z})$ is given by $f(\vec{z}_0)$ for some $\vec{z}_0 \in A \cap \mathbb{E}^+$ where $\mathbb{E}^+ = \{\vec{z} \in \mathbb{E}_4: p_1, p_2, a_1, a_2 > 0\}$. Then the supremum is actually the maximum and we have the nonlinear programming problem: Maximize the objective function $f(\vec{z})$ over $\vec{z} \in \mathbb{E}^+$ subject to the constraints $g_1(\vec{z}) \leq 0$ and $g_2(\vec{z}) \leq 0$ where

$$g_1(\vec{z}) = z_1 z_3^2 + z_2 z_4^2 - P^2,$$

$$g_2(\vec{z}) = z_1 + z_2 - 1.$$

In the terminology of nonlinear programming, \mathbb{E}^+ is the feasible set and any feasible point satisfying constraints $g_1(\vec{z}) \leq 0, g_2(\vec{z}) \leq 0$ is said to be attainable. In the present case, the set of attainable points is exactly the set $A \cap \mathbb{E}^+$. Also in the present case, the feasible set is open and f, g_1, g_2 meet required differentiability conditions [1], so any given attainable point \vec{z}_0 which maximizes f must satisfy the Kuhn-Tucker conditions [1]:

$$\nabla f(\vec{z}_0) - u_1 \nabla g_1(\vec{z}_0) - u_2 \nabla g_2(\vec{z}_0) = 0, \quad (8)$$

$$u_i g_i(\vec{z}_0) = 0, \quad i = 1, 2 \quad (9)$$

$$u_i \geq 0, \quad i = 1, 2$$

where the Lagrange multipliers u_1, u_2 are unique. It follows from (9) that $u_2 = 0$ or $g_2(\vec{z}_0) = 0$. The latter case $g_2(\vec{z}_0) = 0$ reflects the situation in which $p_1 + p_2 = 1$; a situation in which X has only two atoms. If $g_2(\vec{z}_0) \neq 0$ then $u_2 = 0$. Then in (8) we have

$$(a_1 + 1)\ln(a_1 + 1) - a_1 - a_1 \ln(p_1 a_1 + p_2 a_2 + 1) - u_1 a_1^2 = 0, \quad (10)$$

$$(a_2 + 1)\ln(a_2 + 1) - a_2 - a_2 \ln(p_1 a_1 + p_2 a_2 + 1) - u_1 a_2^2 = 0, \quad (11)$$

$$p_1 \ln(a_1 + 1) - p_1 \ln(p_1 a_1 + p_2 a_2 + 1) - 2u_1 p_1 a_1 = 0, \quad (12)$$

$$p_2 \ln(a_2 + 1) - p_2 \ln(p_1 a_1 + p_2 a_2 + 1) - 2u_1 p_2 a_2 = 0. \quad (13)$$

Multiplying (12) by a_1/p_1 and subtracting the result from (10) gives $u_1 = \xi(a_1)$ where

$$\xi(x) = \frac{x - \ln(x+1)}{x^2}, \quad x > 0.$$

A similar calculation using (11) and (13) gives $u_1 = \xi(a_2)$. The inverse ξ^{-1} exists since

$$\xi'(x) = \frac{2}{x^3} \frac{1}{x+1} \left[(x+1)\ln(x+1) - \frac{1}{2}x^2 - x \right] > 0$$

for $x > 0$. u_1 is unique for a given \bar{x}_0 so $a_1 = a_2$, implying that $X \in \mathcal{B}_2(P)$. Thus, in this last case as well, two-level encoding suffices. This proves

Lemma 4:

$$\sup_{X \in \mathcal{B}_3(P)} I[X+1] = \sup_{X \in \mathcal{B}_2(P)} I[X+1].$$

Theorem 1: Suppose the encoder intensity χ_t of a Poisson channel with noise intensity λ is predictable and mean-square-constrained $E[\chi_t^2] \leq P^2$. For this channel

$$\mathcal{C}_{\text{INFO}} = \mathcal{C}_{\text{INFO}}^{\text{OOK}} = D_o(\lambda, P)$$

so that

$$D(0, P) = \frac{2}{e} P$$

and, for $\lambda > 0$,

$$D(\lambda, P) = \max_{a \in A} \frac{P^2}{a} \ln \frac{ak(\lambda/a) + \lambda}{P^2 k(\lambda a/P^2)/a + \lambda} \quad (14)$$

with $A = [a_o, \infty)$ and a_o satisfying $a_o^2 k(\lambda/a_o) = P^2$.

Proof: The case $\lambda=0$ is known [5] to be true. Also in [5], it was shown that $\mathcal{C}_{\text{INFO}}^{\text{OOK}} = D(\lambda, P)$ as given in (14) for $\lambda > 0$. Obviously, $\mathcal{C}_{\text{INFO}} > \mathcal{C}_{\text{INFO}}^{\text{OOK}}$ so all that remains to be shown is that

$$\mathcal{C}_{\text{INFO}} \leq D_o(\lambda, P) \quad (15)$$

for $\lambda > 0$. D and D_o are homogeneous [5] so, actually, it is enough to show (15) for the case $\lambda=1$. From (1) and Jensen's inequality

$$I^T[\theta, Y] \leq I[\chi_t + 1] b([0, T]) = I[\chi_t + 1] T. \quad (16)$$

Here in (16) χ_t is just any random variable belonging to $\mathcal{B}(P)$. Therefore,

$$\mathcal{C}_{\text{INFO}} \leq \sup_{X \in \mathcal{B}(P)} I[X+1] = \sup_{X \in \mathcal{B}_2(P)} I[X+1] \quad (17)$$

where the equality in (17) follows from Lemmas 3 and 4. It follows from [5, Lemma 4] that the RHS of (17) is $D_o(1, P)$. Hence the proof is complete.

CODING CAPACITY

Coding capacity is the threshold on transmission rates below which essentially error-free communication is possible. Coding capacity is founded on ideas of channel codes, decoding schemes, and decoding error probability. A code (M, T, P_e) for the mean-square-constrained Poisson channel is a set of M equally likely nonnegative waveforms $\chi_m(t)$, $t \in [0, T]$, $m = 1, \dots, M$ corresponding to unique messages and satisfying the constraint

$$\frac{1}{M} \sum_{m=1}^M \frac{1}{T} \int_0^T \chi_m^2(t) dt \leq P^2.$$

Let S_Y be the space of trajectories of Y on $[0, T]$. A decoding scheme is a mapping $D: S_Y \rightarrow \{1, 2, \dots, M\}$ where the integers $1, 2, \dots, M$ are labels for members of the set of possible messages. The error probability associated with D is

$$P_e = \frac{1}{M} \sum_{i=1}^M P\{D(Y_0^T) \neq m | \chi_m(\cdot)\}$$

where $Y_0^T \in S_Y$ denotes the path $\{Y_t: t \in [0, T]\}$. A code (M, T, P_e) has rate $R = (1/T) \ln M$. A code rate R is said to be achievable, if for all $\epsilon > 0$, there exists a code (M, T, P_e) whose parameters satisfy $M \geq e^{RT}$ with $P_e \leq \epsilon$ for T sufficiently large. The coding capacity C_{CODING} is the supremum of achievable rates. $C_{\text{CODING}}^{\text{OOK}}$ is the coding capacity of the OOK-constrained Poisson channel in which it is additionally required that the encoded message waveforms $\chi_m(t)$, $m = 1, \dots, M$ of any code (M, T, P_e) take on no more than two values.

Coding capacity can be related to information capacity using Fano's inequality [4].

Lemma 5: $C_{\text{CODING}} \leq \limsup_{T \rightarrow \infty} C_{\text{INFO}}$.

Proof: The proof uses Fano's inequality [2], [4] in a straightforward adaption of proofs [2], [6] given in the context of other channel models.

Applying Lemma 5 to the Poisson channel with and without the OOK constraint, we have

$$C_{\text{CODING}} \leq C_{\text{INFO}}, \quad C_{\text{CODING}}^{\text{OOK}} \leq C_{\text{INFO}}^{\text{OOK}}.$$

We showed in Theorem 1 that $C_{\text{INFO}} = C_{\text{INFO}}^{\text{OOK}}$ and, of course, $C_{\text{CODING}}^{\text{OOK}} \leq C_{\text{CODING}}$ so, to show that the four capacities

$$C_{\text{INFO}}, \quad C_{\text{INFO}}^{\text{OOK}}, \quad C_{\text{CODING}}, \quad C_{\text{CODING}}^{\text{OOK}}$$

are equal, we need only show that $C_{\text{CODING}}^{\text{OOK}} \geq C_{\text{INFO}}^{\text{OOK}}$. This is accomplished using an argument employed by Wyner [10] to address the peak-constrained Poisson channel.

Wyner's approach [10] is to consider the Poisson channel as a binary discrete memoryless channel [6]. The message is assumed to be a stream of 1s and 0s produced at the rate of one symbol θ_n each Δ seconds. The encoded message waveform $\chi_m(t)$ is assumed to be constant in the intervals $((n-1)\Delta, n\Delta]$, $n = 1, 2, \dots$ taking on only the values $\chi_n = 0$ or $\chi_n = a$; 0 if the message symbol is $\theta_n = 0$ and a if the message symbol is $\theta_n = 1$. The receiver decoder identifies the message symbol as 1 if $Y_n \equiv Y_{n\Delta} - Y_{(n-1)\Delta} = 1$. The message symbol is identified as 0 otherwise. The discrete channel created by using the Poisson channel in this way has channel transition probabilities

$$p_{01} = P\{Y_n = 1 | \theta_n = 0\} = \lambda e^{-\lambda \Delta},$$

$$p_{11} = P\{Y_n = 1 | \theta_n = 1\} = (\lambda + a)e^{-(\lambda + a)\Delta}.$$

The coding capacity of the OOK-constrained Poisson channel is lower bounded by the coding capacity of the binary channel with these transition probabilities. Coding capacity and information capacity are equal [6] for the discrete memoryless channel so

$$\Delta C_{\text{CODING}}^{\text{OOK}} \geq \sup_{q \leq \sigma} I[\theta, Y] \quad (18)$$

where $\sigma = P^2/A^2$ and $q = P\{\theta = 1\}$. In (18) and in the definition of q , the subscript n on θ and Y is no longer relevant and has been dropped. The restriction $q \leq \sigma$ in (18) accounts for the mean-square constraint $E[\chi^2] = qA^2 \leq P^2$.

Define $h(x)$ to be the binary entropy function

$$h(x) = -x \ln x - (1-x) \ln(1-x)$$

and let

$$f(q, A) = h(p_{01}(1-q) + p_{11}q) - qh(p_{11}) - (1-q)h(p_{01}).$$

Then

$$\begin{aligned} \sup_{q \leq \sigma} I[\theta, Y] &= \sup_{q \leq \sigma} I[\chi, Y] \\ &= \sup_{q \leq \sigma} [H[Y] - H[Y|\chi]] \\ &= \sup_{q \leq \sigma} [h(p_{01}(1-q) + p_{11}q) - qh(p_{11}) - (1-q)h(p_{01})] \\ &= \sup_{q \leq \sigma} f(q, A). \end{aligned} \quad (19)$$

Let $s = \lambda A$. For small Δ , we can write

$$\begin{aligned} qp_{11} + (1-q)p_{01} &= sA \Delta e^{-sA\Delta}(1-q) + q(s+1)A \Delta e^{-A\Delta(s+1)} \\ &= sA \Delta(1-q)[1 - sA\Delta + o(\Delta)] + q(s+1)A \Delta[1 - A\Delta(s+1) + o(\Delta)] \\ &= sA \Delta(1-q) + qsA \Delta + qA \Delta + o(\Delta) \\ &= A \Delta(s+q) + o(\Delta). \end{aligned}$$

We have $h(x) = -x \ln x + x + O(x^2)$ as $x \rightarrow 0$ so

$$\begin{aligned} f(q, A) &= -(p_{01}(1-q) + p_{11}q) \ln(p_{01}(1-q) + p_{11}q) \\ &\quad + p_{01}(1-q) + p_{11}q + O(p_{01}(1-q) + p_{11}q) \\ &\quad + qp_{11} \ln p_{11} - qp_{11} + O(p_{11}^2) \\ &\quad + (1-q)p_{01} \ln p_{01} - (1-q)p_{01} + O(p_{01}^2) \\ &= -(p_{01}(1-q) + p_{11}q) \ln(p_{01}(1-q) + p_{11}q) \\ &\quad + qp_{11} \ln p_{11} + (1-q)p_{01} \ln p_{01} + o(\Delta) \\ &= (1-q)sA \Delta \ln(sA \Delta) + q(s+1)A \Delta \ln((s+1)A \Delta) \\ &\quad - (s+q)A \Delta \ln((s+q)A \Delta) + o(\Delta) \\ &= A \Delta \left[(1-q)s \ln s + (1-q)s \ln(A \Delta) \right. \\ &\quad \left. + q(s+1) \ln(s+1) + q(s+1) \ln(A \Delta) \right. \\ &\quad \left. - (s+q) \ln(s+q) - (s+q) \ln(s+q) \ln(A \Delta) \right] + o(\Delta) \\ &= A \Delta \left[(1-q)s \ln s + q(s+1) \ln(s+1) - (s+q) \ln(s+q) \right] + o(\Delta). \end{aligned} \quad (20)$$

When combined, (18), (19) and (20) give

$$\mathbb{C}_{\text{CODING}}^{\text{OOK}} \geq \sup_{q \leq \sigma} A \left[(1-q)s \ln s + q(s+1) \ln(s+1) - (s+q) \ln(s+q) \right].$$

We have

$$(1-q)s \ln s + q(s+1) \ln(s+1) = \frac{1}{A} \left[(1-q)\lambda \ln \lambda + q(\lambda+A) \ln(\lambda+A) \right] - \frac{1}{A} (\lambda+qA) \ln A$$

and

$$(s+q) \ln(s+q) = \frac{1}{A} (\lambda+qA) \ln(\lambda+qA) - \frac{1}{A} (\lambda+qA) \ln A.$$

Therefore

$$\begin{aligned} \mathbb{C}_{\text{CODING}}^{\text{OOK}} &\geq \sup_{q \leq \sigma} \left[(1-q)\lambda \ln \lambda + q(\lambda+A) \ln(\lambda+A) - (\lambda+qA) \ln(\lambda+qA) \right] \\ &= \sup_{X \in \mathcal{B}_2(P)} \left[E[(X+\lambda) \ln(X+\lambda)] - (E[X]+\lambda) \ln(E[X]+\lambda) \right] \\ &= \sup_{X \in \mathcal{B}_2(P)} I[X+\lambda] \\ &= \mathbb{C}_{\text{INFO}}^{\text{OOK}}. \end{aligned}$$

This proves

Theorem 2: For the Poisson channel with mean-square-constrained encoder intensity $E[\chi_i^2] \leq P^2$ and noise intensity λ ,

$$\mathbb{C}_{\text{INFO}} = \mathbb{C}_{\text{INFO}}^{\text{OOK}} = \mathbb{C}_{\text{CODING}} = \mathbb{C}_{\text{CODING}}^{\text{OOK}} = D(\lambda, P).$$

The capacities in Theorem 2 are all capacities of the Poisson channel with causal feedback. However, as observed in [5], \mathbb{C}_{INFO} and $\mathbb{C}_{\text{INFO}}^{\text{OOK}}$ are also no-feedback capacities. Therefore $\mathbb{C}_{\text{CODING}}$ and $\mathbb{C}_{\text{CODING}}^{\text{OOK}}$ are also no-feedback capacities.

OTHER REMARKS

We are justified to an extent by Theorem 2 in speaking of the capacity of the mean-square-constrained Poisson channel without using qualifying descriptors; for nonrandom noise intensity λ all the capacities - information and coding, with and without OOK constraint, with and without causal feedback - are the same. The availability of the expression (14) for the channel capacity enables us to consider questions of time-varying channel parameters $\lambda(t)$ and $P(t)$. Strictly analogous to proof for the peak-constrained Poisson channel in [3], it can be shown that for b -measurable $\lambda(t)$ and $P(t)$

$$C_{\text{INFO}} = \frac{1}{T} \int_0^T D(\lambda(t), P(t)) b(dt). \quad (21)$$

It is easily shown that (21) is the information capacity with or without causal feedback. Further restrictions on $\lambda(t)$ and $P(t)$ to provide for stationarity such as periodicity or almost periodicity allow expressions analogous to (21) to be given for C_{CODING} . See [4].

Given (21) for the information capacity, it is possible to give a treatment of jamming along the lines given in [3] for the peak-constrained Poisson channel. We take as our jamming model

$$\eta_t = \chi_t + \lambda(t) + J_t$$

where η_t is the intensity of the channel output and J_t is the intensity - possibly stochastic - of the Poisson-type process G_t jamming the channel. Also, we make the assumptions that G_t is independent of the message and noise processes θ_t and N_t and that

$$\frac{1}{T} \int_0^T E[J_t] b(dt) \leq P_J.$$

Under these assumptions it can be shown that from the standpoint of minimizing information capacity, the optimal jamming intensity is nonrandom and follows the form of a waterfilling scheme [2]; i.e.,

$$J_t^{\text{opt}} = [\alpha P(t) - \lambda(t)]^+$$

where $\alpha \geq 0$ satisfies

$$\frac{1}{T} \int_0^T [\alpha P(t) - \lambda(t)]^+ b(dt) = P_J.$$

It is hoped that knowledge of the form of $D(\lambda, P)$ given here will also shed light on other questions relating to channel error exponent, random noise intensity, and marked Poisson channels.

REFERENCES

- [1] M.S. Bazaraa and C.M. Shetty, *Nonlinear Programming, Theory and Algorithms*, John Wiley & Sons, New York, 1979.
- [2] R. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley Publ. Co., Reading, MA., 1987.
- [3] M.R. Frey, "Capacity of the Poisson Channel with Time-Varying Noise Intensity and Jamming," *LISS Tech. Report Series*, No. 25, Department of Statistics, University of North Carolina, May 1988.
- [4] M.R. Frey, "Coding Capacity and Error Exponent of the Poisson Channel with Time-Varying Channel Parameters," *LISS Tech. Report Series*, No. 38, Department of Statistics, University of North Carolina, August 1989.
- [5] M.R. Frey, "Information Capacity of the Poisson Channel with Mean-Square-Constrained Encoder Intensity," *LISS Tech. Report Series*, No. 43, Department of Statistics, University of North Carolina, April 1990.
- [6] R.G. Gallager, *Information Theory and Reliable Communication*, Wiley & Sons, Inc., New York, 1968.
- [7] W. Hoeffding, "The Extrema of the Expected Value of a Function of Independent Random Variables," *Annals of Mathematical Statistics*, Vol. 26, pp. 268-275, 1955.
- [8] R. Liptser and A. Shiriyayev, *Statistics of Random Processes, Vol. II, Applications*, Springer-Verlag, New York, 1977.
- [9] M.S. Pinsker, *Information and Information Stability of Random Variables and Processes*, Holden-Day, San Francisco, 1964.
- [10] A.D. Wyner, "Capacity and Error Exponent for the Direct Detection Photon Channel - Part I, II," *IEEE Trans. on Info. Theory*, Vol. 34, No. 6, November 1988.